

# 2nd January 2020

## Categorizing Humanities clusters

Using their ERIH PLUS discipline

crossref

## ERIH disciplines

Only from main AH journals

For each cluster retrieve the ERIH PLUS disciplines

	Cluster ID	Discipline name	Discipline pubs	ERIH pubs	Cluster pubs	Discipline/ERIH	Discipline/Cluster	score<Perc AH pubs>
1	105345	Religious Studies and Theology	3122	3613	19942	0.864101854414614	0.156554006619196	0.1811754
2	105345	Philosophy	3430	3613	19942	0.949349570993634	0.171998796509879	0.1811754
3	1089010	Literature	13	63	441	0.206349206349206	0.0294784580498866	0.1428571
4	1089010	History	15	63	441	0.238095238095238	0.0340136054421769	0.1428571
5	1089010	History & Philosophy of Science	20	63	441	0.317460317460317	0.0453514739229025	0.1428571
6	1089010	Philosophy	30	63	441	0.476190476190476	0.0680272108843537	0.1428571
7	11076962	History & Philosophy of Science	48	171	234	0.280701754385965	0.205128205128205	0.7307692
8	11076962	Philosophy	130	171	234	0.760233918128655	0.555555555555556	0.7307692
9	11088013	History	45	126	1943	0.357142857142857	0.0231600617601647	0.06484817
10	11088013	Political Sciences and International Relations	53	126	1943	0.420634920634921	0.0272774060730829	0.06484817
11	11126184	Religious Studies and Theology	93	117	135	0.794871794871795	0.688888888888889	0.8666667
12	11128814	Philosophy	136	169	385	0.804733727810651	0.353246753246753	0.438961
13	11137739	Anthropology	8	36	351	0.222222222222222	0.0227920227920228	0.1025641
14	11137739	Political Sciences and International Relations	8	36	351	0.222222222222222	0.0227920227920228	0.1025641
15	11137739	Interdisciplinary research in the Social Scie...	9	36	351	0.25	0.0256410256410256	0.1025641
16	11137739	History	20	36	351	0.555555555555556	0.056980056980057	0.1025641
17	11179450	History	1387	6874	13914	0.201774803607798	0.0996837717406928	0.4940348
18	11179450	Anthropology	1667	6874	13914	0.242508001163806	0.119807388242058	0.4940348
19	11179450	Archaeology	6805	6874	13914	0.989962176316555	0.489075751042116	0.4940348
20	11199213	Linguistics	1137	1247	2662	0.911788291900561	0.427122464312547	0.4684448
21	11222457	Musiology	222	239	350	0.928870292887029	0.634285714285714	0.6828572
22	11225843	History	39	44	190	0.886363636363636	0.205263157894737	0.2315789
23	11235407	History	38	65	236	0.584615384615385	0.161016949152542	0.2754237

**which ratio is more significant to display? The number of pubs for discipline/number of total AH pubs from ERIH or over the total number of publications within the Cluster?**

## Solution with only main disciplines

**N.B.** The discipline is based on journals and a single journal can have multiple disciplines. This means that a single pub can be treated in more disciplines.

id	label	x	y	cluster	weight<Links>	weight<Total link strength>	weight	score<Perc AH pubs>	score<Perc DH pubs>	score<Perc DH pubs level 1>	score<Perc DH pubs level 2>	score<Perc DH pubs level 3>	description	Main Discipline	Discipline pubs	
1	31312285	4022	-0.3832	0.8048	1	2	1	50	0.98	0	0	0	table<tr><td>Micro-level field</td><td>4023</td></tr>	History	49	
2	210130	394	-0.3435	0.74	1	535	3.319	6527	0.904207	0.003320341	0	0.001154901	0.00216544	table<tr><td>Micro-level field</td><td>395</td></tr>	Philosophy	6084
3	1126184	3624	-0.3940	0.8511	1	28	1.0061	135	0.866867	0	0	0	table<tr><td>Micro-level field</td><td>3625</td></tr>	Religious Studies and Theology	93	
4	31338307	4014	-0.4759	0.7305	1	10	1.0015	51	0.8627451	0	0	0	table<tr><td>Micro-level field</td><td>4015</td></tr>	Religious Studies and Theology	43	
5	20170263	2435	-0.1905	0.5018	1	123	1.0672	652	0.8526012	0.005780347	0.002890173	0	0.002890173	table<tr><td>Micro-level field</td><td>2436</td></tr>	Archaeology	551
6	1874230	1237	-0.4037	0.7072	1	406	1.6349	2508	0.84889	0.001594896	0	0.001594896	0	table<tr><td>Micro-level field</td><td>1238</td></tr>	Philosophy	2958
7	22512336	2203	-0.446	0.7653	1	116	1.3002	892	0.8396861	0.05304933	0.0190583	0.05453274	0.05453274	table<tr><td>Micro-level field</td><td>2204</td></tr>	Linguistics	737
8	1862356	810	-0.4301	0.5997	1	212	1.6096	4033	0.8265278	0.00777307	0	0.00440179	0.00310109	table<tr><td>Micro-level field</td><td>811</td></tr>	Linguistics	3587
9	11355737	552	0.0205	0.679	1	810	3.1543	5510	0.8150563	0.0005967441	0	0.0005444646	0.0005444646	table<tr><td>Micro-level field</td><td>553</td></tr>	Philosophy	4367
10	31339560	3462	-0.4268	0.7325	1	34	1.0097	188	0.7872341	0	0	0	0	table<tr><td>Micro-level field</td><td>3463</td></tr>	Religious Studies and Theology	112
11	31351301	3994	-0.4007	0.8638	1	26	1.0029	54	0.7777778	0	0	0	0	table<tr><td>Micro-level field</td><td>3995</td></tr>	History	38
12	19528165	2050	-0.4387	0.6988	1	133	1.1505	1037	0.7608486	0.00293153	0	0.005785921	0.07714561	table<tr><td>Micro-level field</td><td>2051</td></tr>	Linguistics	773
13	31312009	4001	0.0196	0.6794	1	14	1.0029	53	0.754717	0	0	0	0	table<tr><td>Micro-level field</td><td>4002</td></tr>	Philosophy	38
14	1901012	1352	1.1852	0.9179	5	182	2.723	2134	0.7497657	0	0	0	0	table<tr><td>Micro-level field</td><td>1353</td></tr>	Philosophy	1564
15	17473088	2038	-0.2759	0.6763	1	252	1.2658	1053	0.7350428	0.004748338	0	0	0.004748338	table<tr><td>Micro-level field</td><td>2039</td></tr>	Archaeology	638
16	11070962	3322	-0.369	0.785	1	115	1.0622	234	0.7307692	0	0	0	0	table<tr><td>Micro-level field</td><td>3323</td></tr>	Philosophy	130
17	14012219	1206	-0.4391	0.8134	1	374	1.4017	2610	0.7133908	0.04099617	0.00051341	0.006396552	0.02758621	table<tr><td>Micro-level field</td><td>1207</td></tr>	Linguistics	1717
18	17786971	3444	-0.4478	0.8238	1	70	1.0026	195	0.7460817	0.122449	0	0	0.122449	table<tr><td>Micro-level field</td><td>3445</td></tr>	Literature	121
19	27167592	1626	-0.3237	0.7435	1	297	1.3136	1326	0.7016199	0.004024087	0.00202444	0	0.00202444	table<tr><td>Micro-level field</td><td>1627</td></tr>	Archaeology	650
20	17647489	1032	-0.3964	0.8852	1	507	3.802	3115	0.6959872	0.01348315	0.0006420546	0	0.01204109	table<tr><td>Micro-level field</td><td>1033</td></tr>	History	1897
21	17745675	3334	-0.4618	0.6894	1	114	1.039	230	0.6913943	0	0	0	0	table<tr><td>Micro-level field</td><td>3335</td></tr>	History	149

Working on displaying the graph.

- Gephi: Too much data. It crashes.
- VosViewer: Can't manage other kinds of colors
- Cytoscape: Layout in progress

## Digital Humanities and Humanities

id	label	x	y	cluster	weight<Links>	weight<Total link strength>	weight	score<Perc AH pubs>	score<Perc DH pubs>	score<Perc DH pubs level 1>	score<Perc DH pubs level 2>	description	Main Discipline	Discipline pubs	
1	1880760	2409	-0.2336	0.7465	1	307	1.3318	718	0.4164345	0.6657382	0.270195	0.3955432	table<tr><td>Micro-level field</td><td>2410</td></tr>	Linguistics	263
2	1866507	3230	-0.3385	0.7744	1	82	1.0314	268	0.608209	0.5410448	0.4962687	0.04477612	table<tr><td>Micro-level field</td><td>3231</td></tr>	Linguistics	1111
3	19513346	1004	0.2808	0.5548	5	613	2.6757	3219	0.3603604	0.3470022	0.01988195	0.3271202	table<tr><td>Micro-level field</td><td>1005</td></tr>	Linguistics	1136
4	1704728	3646	0.5613	0.4038	5	90	1.0439	129	0	0.07751938	0	0.07751938	table<tr><td>Micro-level field</td><td>3647</td></tr>	NULL	NULL
5	1430295	1138	0.1912	0.5972	5	846	2.5371	2791	0.1640989	0.05589395	0.000716589	0.05517736	table<tr><td>Micro-level field</td><td>1139</td></tr>	Philosophy	358
6	254406	637	-0.2223	0.7429	1	889	2.43	4908	0.03198859	0.05093725	0.003463733	0.04747351	table<tr><td>Micro-level field</td><td>638</td></tr>	NULL	NULL
7	31315543	3607	-0.1544	0.5924	1	163	1.0493	142	0.09859155	0.04225352	0	0.04225352	table<tr><td>Micro-level field</td><td>3608</td></tr>	History	3
8	31315543	3607	-0.1544	0.5924	1	163	1.0493	142	0.09859155	0.04225352	0	0.04225352	table<tr><td>Micro-level field</td><td>3608</td></tr>	Musicology	3
9	22512336	2203	-0.446	0.7653	1	116	1.3002	892	0.8396861	0.03811659	0.0190583	0.0190583	table<tr><td>Micro-level field</td><td>2204</td></tr>	Linguistics	737
10	1863817	2254	0.5677	0.3962	5	354	1.4092	844	0.01658768	0.03436019	0.02014218	0.01421801	table<tr><td>Micro-level field</td><td>2255</td></tr>	NULL	NULL
11	1874100	1255	-0.4046	0.705	1	809	2.3245	2459	0.4221228	0.03416023	0	0.03416023	table<tr><td>Micro-level field</td><td>1256</td></tr>	Philosophy	954

The top 10 main DH clusters: (in crossref the DH clusters are related to Linguistic)

ID	weight	Perc AH pubs	Perc DH pubs	Main Discipline	Discipline pubs	Most frequent terms
2409	718	0.4164345	0.6657382	Linguistics	263	authorship; analysis; attribution; using; language; text; study; classification; based; approach;
3230	268	0.608209	0.5410448	Linguistics	111	english; digital; tei; early; analysis; case; texts; text; edition; scholarly;
1004	3219	0.3603604	0.3470022	Linguistics	1136	analysis; sentiment; using; translation; language; machine; semantic; text; based; approach;

3646	129	0	0.07751938	NULL	NULL	NULL
1138	2791	0.1640989	0.05589395	Philosophy	358	NULL
637	4908	0.03198859	0.05093725	NULL	NULL	NULL
3607	142	0.09859155	0.04225352	History	3	NULL
3607	142	0.09859155	0.04225352	Musicology	3	NULL
2203	892	0.8396861	0.03811659	Linguistics	737	NULL
2264	844	0.01658768	0.03436019	NULL	NULL	
1255	2459	0.4221228	0.03416023	Philosophy	954	

### parameters:

I've decided to use these parameters now for having only the important elements

```
-- % AH pubs
declare @AH_pubs float = 0.05

-- % discipline / cluster num of pub
declare @Discipline_cluster float = 0.01

-- % discipline / num of AH pub
declare @Discipline_ERIH float = 0.2

drop table #data_filter
select *
into #data_filter
from #data
where [Discipline/ERIH] >= @Discipline_ERIH
and [Discipline/Cluster] >= @Discipline_cluster
and [score<Perc AH pubs>] >= @AH_pubs
order by [Cluster ID], [Discipline/ERIH]
```